

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

Remarks

This Response is responsive to the Final Office Action of **January 30, 2006**.
Reexamination and reconsideration of **claims 1 and 4-21** is respectfully requested.

Summary of The Final Office Action

Claims 1, 4-13 were rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy (US 6,125,362).

Claims 15, 16, 20, and 21 were rejected under 35 U.S.C. §102(b) as being anticipated by Pon et al. (US 6,047,251).

Claims 14, 17-19 were rejected under 35 U.S.C. §103(a) as being unpatentable over Pon in view of Elworthy.

The Present Claims Patentably Distinguish Over the References of Record

Independent Claim 1

Claim 1 was rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy (US 6,125,362). Applicant respectfully submits that Elworthy does not teach each and every element of claim 1 and thus fails to support the §102 rejection. Therefore, the rejection should be withdrawn.

For example, the Office Action cites Elworthy (column 7, lines 50-65) as teaching the claimed database and the claimed limitation of:

"each text string of the plurality of text strings having an associated probability value indicating a probability that the text string occurs within a language based on occurrences of the text string in all of the candidate languages"

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

The cited section of Elworthy, column 7, lines 50-65 is reproduced here as:

"Methods which can be used are the methods described in the articles by Sibun & Spitz and Sibun & Reynar. The word tokens are then input to each of the lexicons 25a,25b,25c . . . 25L for the languages to which the OCR data may belong. The lexicons 25a,25b,25c . . . 25L comprise **predetermined probability values that the word token belongs to the language.** The probability output from the lexicons 25a,25b,25c . . . 25L are input to respective accumulators 26a,26b,26c . . . 26L where the probabilities for sequential word tokens are accumulated to form an accumulated probability. The accumulated probabilities of each of the accumulators 26a,26b,26c . . . 26L are input to a comparator 26 wherein the probabilities are compared with one another and with a predetermined threshold to determine whether a language is uniquely identifiable as the language to which the OCR data belongs." (emphasis added)

From this passage, Elworthy teaches that the lexicons comprise predetermined probability values that the word token belongs to the language. Thus, there is no teaching or suggestion that the probability is "based on occurrences of the text string in all of the candidate languages" as claimed. In fact, Elworthy states that probabilities are "independent" between languages:

"...the probabilistic model for one language is **independent of the others.**" (Elworthy, column 8, lines 5-7) [emphasis added].

This is consistent with the technique of Elworthy, which accumulates probabilities (e.g. adds total points) for each language. Therefore, Elworthy fails to teach the claimed database and the claimed probability values as recited in claim 1. For at least this reason, the rejection is not supported by Elworthy and should be withdrawn.

To further show that the recited database and its claimed limitations differ from Elworthy's lexicons 25a-L, the following is provided. Applicant respectfully submits that the presently claimed system provides for an elegant and efficient technique for determining a document language. With the present system, the probabilities used can be implemented with, for example, simple calculations like that shown by equation (1) on

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

page 7 of the present specification. Equation (1) shows an example of a probability of a string being in English that is "based on occurrences of the text string in all of the candidate languages". Conversely, Elworthy seems to involve a complex technique that uses many complicated formulas. For example, Elworthy explains how the probability values of its training data are calculated and uses 4 columns (columns 8-11) to describe the lengthy process. The process includes processing at least 17 equations until the training procedure is terminated (column 11, lines 31-37, see equations 1-17). Understanding these differences may assist in determining that the present claims recite different systems and methods than disclosed by Elworthy.

With reference to the recited "language analyzer" of claim 1, the language analyzer retrieves probability values from the database and adjusts a negative assumption until a language is determined. This features is also not disclosed by Elworthy.

Elworthy determines the language of a document by accumulating probabilities for each token in each language for a set of input data and then compares the final accumulated total to see which language has the highest total based on a threshold (see Elworthy, Figure 5, Figure 14a-c, column 13, lines 36-58). Thus, Elworthy simply counts up scores until the input data is processed and compares the total scores. This is basically a type of a "first man to the finish line" test where all candidates run the entire race to see if there is a clear winner. There is no "proof" or "proving or disproving assumptions" since it simply counts points to see who has the most points.

The present system is different. A negative assumption is set and it is adjusted based on the recited predetermined probabilities (which are different from Elworthy). The claimed system is more like a "last man standing" analysis rather than a "first man to the finish line." For example, the present specification shows example results in Table 1 on page 12 where after only 2 iterations, the candidate language of English can be eliminated as a possibility. Therefore, the problem space in the present system can be reduced whereas the problem space of Elworthy is linear (e.g. process all input data to

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

determine who finishes the race). Furthermore, Elworthy explains that more than one language may cross the threshold and be a possible winner (column 13, lines 49-54). This further shows that Elworthy does not "prove" or "disprove" assumptions but simply counts and accumulates scores. Crossing the threshold in Elworthy does not prove a language as acknowledged by Elworthy (column 13, lines 49-54).

Since claim 1 recites features not taught or suggest by Elworthy, Elworthy fails to support the §102 rejection and the rejection must be withdrawn. As such, claim 1 patentably distinguishes over the references of record and is in condition for allowance. Accordingly, dependent claims 4-6 also patentably distinguish over the references and are in condition for allowance.

Independent Claim 7

Claim 7 was also rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy. Applicant respectfully submits that Elworthy does not teach each and every element of claim 7 and thus fails to support the §102 rejection. Therefore, the rejection should be withdrawn.

In particular, with reference to the recited element of "setting a null hypothesis to a true value...", the Office Action cites Elworthy column 12, lines 20-38 and claim 13. In this section, Elworthy states, "the accumulator is initially zeroed..." (column 12, line 22). Then the accumulator adds probabilities to a total value and when finished, the totals are compared. As previously explained, this is simply a counting technique that starts at zero. Setting the accumulator to zero does not teach or equate to setting a null hypothesis to a true value. The analysis is different. Applicant understands the Examiner's interpretation of this position but in the mathematical arts, one of ordinary skill would not equate these features to be the same, neither in purpose nor function.

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

Furthermore, it is not the intent or purpose of Elworthy to set a null hypothesis and then disprove the null hypothesis as in claim 7. Elworthy tries to determine a language of a document in a different way, namely, accumulating probability values. Applicant invites the Examiner to review paragraphs [0036] to [0045] of the present specification for a better understanding of how the present method may operate.

With reference to the recited element of "determining a contrary probability...", Elworthy fails to teach each and every associated limitation as recited in claim 7. As previously explained, Elworthy determines probabilities of languages "independent of the others." (Elworthy, column 8, lines 5-7). Thus, Elworthy does not disclose determining a contrary probability "based on probabilities that the text string belongs to each of the candidate languages" as recited in claim 7. Elworthy further does not disclose the recited element of "based on occurrences of the text string in all of the candidate languages." Elworthy is not based on this type of probability model and thus fails to teach or suggest these features of claim 7. Thus for at least this reason, Elworthy fails to support the §102 rejection and the rejection should be withdrawn.

Regarding the limitation of "determining the document is one language...", the Office Action cites Elworthy column 12, lines 20-38 and column 13, lines 22-35, and reasons that "the highest accumulated probability-accounts for approval and simultaneously disapproval, C. 13, lines 44-58." Although this may make some sense upon first glance, it is actually not accurate. As previously explained, Elworthy accumulates probabilities from an initial starting point of zero (column 12, line 22). However, when accumulation is finished, there may not be what the Office Action states is an approval or simultaneous disapproval. That is because Elworthy does not "prove" or "disprove" anything but rather simply counts points. But even when the points are counted and the threshold is passed, there may not be a winner and nothing is proved as Elworthy explains:

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

"In FIG. 14b it can be seen that the probability of the language being English has exceeded the threshold but there is still overlap with the probability for the languages being French and Italian. If there is no more data these **three languages could be identified as possible languages** to which the input data belongs." (Elworthy, column 13, lines 49-54, and Figure 14b) [emphasis added]

Thus, passing the threshold does not prove a language and does not simultaneously disprove an opposite hypothesis, and this is acknowledged by Elworthy. Indeed, Elworthy simply starts a counter at zero and accumulates probabilities (column 12, line 22, "the accumulator is initially zeroed"). Therefore, Elworthy does not teach the claimed "determining" element that includes disproving the null hypothesis by approaching the false value. For at least this reason, a proper §102 rejection has not been established and the rejection should be withdrawn.

Based on the above explanations, Elworthy fails to teach each and every feature of claim 7. Thus, Elworthy fails to support a proper §102 rejection and the rejection must be withdrawn. As such, claim 7 patentably distinguishes over the references of record and is in condition for allowance. Accordingly, dependent claims 8-14 also patentably distinguish over the references and are in condition for allowance.

Dependent Claim 10

Claim 10 recites pregenerating probability data corresponding to each candidate language, the probability data including a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages. The Office Action (page 6) cites Elworthy column 2, lines 30-38, which discusses comparing probability values, as teaching normalization.

Applicant respectfully submits that normalizing a probability value as recited in claim 10 involves and results in changing the values of the probability values. This is understood by one of ordinary skill in the art. Simply "comparing" values as Elworthy performs does not change values and is a different process. Indeed, one of ordinary skill

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

in the art understands that "normalizing" is not "comparing" and that "comparing" does not teach "normalizing".

Elworthy does not mention normalizing or any form of normalization in its disclosure. That is because Elworthy does not involve normalization. As previously explained, "the probabilistic model for one language is independent of the others." (Elworthy, column 8, lines 5-7). This is contrary to normalization techniques and is contrary to the present method. Thus the method of claim 10 is not taught or suggested by Elworthy and the rejection should be withdrawn.

Independent Claim 15

Claim 15 was rejected under 35 U.S.C. §102(b) as being anticipated by Pon. Applicant respectfully submits that Pon does not teach each and every element of claim 15 and thus fails to support the §102 rejection. Therefore, the rejection should be withdrawn.

Pon discloses optical character recognition system that uses a dictionary-based approach to identify languages in a document (see Abstract). Pon uses a technique that is basically a stripped-down version of Elworthy. Instead of accumulating complex probabilities that Elworthy generates, Pon simply counts the number words from a document that matches a dictionary for a specific language.

"the confidence statistic can be computed by counting the number of words in the zone that are found in each of the respective dictionaries." (Pon, column 5, lines 63-65).

"The language with the highest confidence statistic is ascertained, and used as an initial estimate of the language for the zone." (Pon, column 6, lines 1-3).

The "confidence statistic" as described by Pon is simply an accumulated total of the number of words in a document region that are found in a dictionary. Basically, Pon adds a "1" to an counter for each word match or "0" for no match (Pon, column 7, lines

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

1-3). When processing is finished, Pon compares the total points counted for each language and looks for the highest score (Pon, column 8, lines 4-9). This appears to be the same analysis technique as used by Elworthy with the main difference being that Elworthy generates its own lexicons while Pon uses dictionaries.

Similar to the above explanation that Elworthy does not "prove" or "disprove" anything but rather is simply accumulating points, Pon also does not "prove" or "disprove" any assumptions. Pon simply counts dictionary matches.

Applicant believes to understand the Examiner's interpretation of Pon. For example, when Pon resets its counter to "0" prior to counting/accumulating matches (column 7, line 37), the Examiner reads this on setting an assumption and then when words are matched, the counter is incremented, thus disproving the assumption. Applicant respectfully submits that Pon is a different type of analysis than the present process of claim 15 and one of ordinary skill in the art would not understand or interpret Pon in this manner. One of ordinary skill would not equate counting word matches from dictionaries to teach the recited setting and disproving probability assumptions.

Further as the Examiner states, when Pon determines that a word belongs to a language, it inherently disproves that the word does not belong to the language. However, what Pon fails to teach is any technique that relates to the disproving. For example in Pon, a "1" proves that a word is in a dictionary and a "0" (the contrary) proves that the word is not. Pon accumulates all values of "1" for each language to obtain a total confidence statistic (e.g. total score) (column 8, lines 4-9).

However, to teach the claimed disproving features of claim 15 based on the Examiner's interpretation, Pon must teach a process that uses the "0" values (contrary probabilities) to determine the language of a document. Of course, since Pon is a counting technique, it does nothing with the "0" values. Zeros are discarded. Thus, the "0" values do not prove or disprove any assumption, and no process is disclosed that uses

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

the "0" values to determine the language of a document. Therefore, Pon fails to teach or suggest setting a probability assumption indicating that the document is not in the selected language and disproving the probability assumption based on a contrary probability as recited in claim 15.

Claim 15 also recites "if the contrary probability fails to support the probability assumption, then the document is determined as being in the selected language." Applicant finds no teaching in Pon where the "0" values are used to perform this process.

Lastly, let's assume the Examiner's reasoning is correct that when a reference proves an assumption, then it inherently disproves the contrary. However, the Applicant respectfully submits that more is needed to reject the present claims. Claim 15 recites a particular process for disproving a probability assumption and this claimed process and its recited limitations are not taught by the references, individually or in combination.

For the reasons set forth above, a proper §102 rejection of claim 15 has not been established since Pon fails to teach each and every limitation recited in claim 15. The rejection should therefore be withdrawn. As such, claim 15 is now in condition for allowance. Accordingly, dependent claims 16-21 are also in condition for allowance.

103 Rejections

Claims 14 and 17-19 were rejected under 35 U.S.C. 103(a) as being unpatentable over Pon in view of Elworthy (US 6,125,362). Applicant respectfully submits that Elworthy fails to teach or suggest the claimed features. In particular, Elworthy teaches independent probabilities: "the probabilistic model for one language is independent of the others." (column 8, lines 5-7). Thus, Elworthy fails to teach or suggest the feature relating to normalization as recited in claims 14 or 19, or "where the contrary probability of a character string in one language is determined based on an occurrence frequency of the character string in the one language influenced by a total occurrence frequency of the

Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

character string in all the candidate languages" as recited in claim 17. Thus Elworthy fails to cure the shortcomings of Pon.

A more detailed explanation of why Elworthy fails to teach or suggest normalization techniques is provided with reference to claim 10 above.

On page 9 of the Office Action, citation is made to Elworthy's discussion of the variable "M", which is the frequency of a word in all languages (Elworthy, column 10, line 18) and the alleged interpretation that "p(m)" is a normalization. Applicant submits that this interpretation is not accurate. The variable "m" is the frequency of a word in a language (Elworthy, column 10, line 17). Therefore, to obtain a normalized frequency of a word in all languages, the calculation would involve solving m/M . This calculation is not performed by Elworthy and is not involved in the disclosure of Elworthy. Thus, Elworthy fails to teach or suggest the alleged normalization and thus fails to teach or suggest the recited limitations relating to normalizing in present claims 10, 14 and 19.

For these additional reasons, claims 14 and 17-19 patentably distinguish over the references of record and are in condition for allowance.


Appl. No. 09/884,403
Response Dated March 30, 2006
Reply to Office Action of January 30, 2006

Conclusion

For the reasons set forth above, **claims 1 and 4-21** patentably and unobviously distinguish over the references of record and are now in condition for allowance. An early allowance of all claims is earnestly solicited.

Respectfully submitted,

MARCH 30, 2006


PETER KRAGULJAC (Reg. No. 38,520)

(216) 348-5843

McDonald Hopkins Co., LPA
600 Superior Avenue, E.
Suite 2100
Cleveland, OH 44114